16S Ribosomal DNA Sequence-Based Analysis of Clinically Significant Gram-Positive Anaerobic Cocci

Yuli Song,[1]* Chengxu Liu,[1] Maureen McTeague,[2] and Sydney M. Finegold[3,4,5]

Research Service,[1] Clinical Microbiology Laboratory,[2] and Infectious Diseases Section,[3] VA Medical Center West Los Angeles, and Department of Medicine[4] and Department of Microbiology, Immunology, and Molecular Genetics,[5] University of California at Los Angeles School of Medicine, Los Angeles, California 90073

Sequence analysis of the 16S rRNA gene represents a highly accurate and versatile method for bacterial classification and identification, even when the species in question is notoriously difficult to identify by phenotypic means. However, its use for identification based on public sequence databases is not without limitation due to the presence of ambiguous data in the databases. In this study, we evaluated the utility of 16S ribosomal DNA sequencing as a means of identifying clinically important gram-positive anaerobic cocci (GPAC) by sequencing 13 type strains of established GPAC species and 156 clinical isolates that had been studied only by phenotypic tests. Among the 13 type strains of GPAC species we tested, only 4 gave a "perfect" match with their corresponding sequences in GenBank, whereas the other 9 had lower sequence similarities (<98%). This indicates that data in the public database may be inaccurate at times. Based on the sequences of the 13 type strains obtained in this study, 84% (131 of 156) of the clinical isolates were accurately identified to species level, with the remaining 25 clinical strains revealing nine unique sequences that may represent eight novel species. This finding is in contrast to the phenotypic identification results, by which only 56% of isolates were correctly identified to species level.

Gram-positive anaerobic cocci (GPAC) are part of the commensal flora in human and animals and are also commonly associated with a variety of human infections (10, 15); data from four surveys of anaerobic infections (2, 14, 19, 37) are consistent in indicating that they account for about 25 to 30% of all anaerobic isolates. Extensive taxonomic changes have occurred recently among this group of bacteria (6, 20), especially in clinically important genera such as *Finegoldia*, *Micromonas*, *Peptostreptococcus*, and *Peptoniphilus*. Although these bacteria have been cultured from a wide variety of infections of the mouth, skin and soft tissues, bone and joints, and upper respiratory and female genital tracts (10, 11, 15), the clinical importance of individual GPAC strains is poorly understood due to the lack of a sound classification and identification scheme.

Early identification schemes of GPAC depended on the microscopic appearance, colonial morphology, and carbohydrate fermentation reactions (28). However, these tests proved to be of limited value for this group of bacteria that are often pleomorphic and usually asaccharolytic. Gas-liquid chromatography was introduced for the detection of volatile fatty acid end products of metabolism, but most GPAC produce a very limited range of volatile fatty acids. In the 1980s, Ezaki et al. (7) found that proteolytic enzyme profiles could distinguish clearly and reproducibly among recognized species of GPAC; this contributed to the development of several commercial preformed enzyme kits, such as RapID ANA (Innovative Diagnostic Systems, Atlanta, Ga.) and Rapid ID 32A (API-bio-

Merieux, Basingstoke, United Kingdom). These commercial kits represent a considerable advance in identification methods in terms of speed, simplicity, and discrimination (16, 23). However, these kits are designed to identify as wide a range of anaerobes as possible, and they contain many tests of little relevance for the identification of GPAC. Furthermore, databases accompanying the kits are often incomplete or inaccurate, especially with a rapid increase of newly described species. The challenge now is to develop a more reliable classification and identification scheme, so that most clinical strains can be allocated to clearly defined, phylogenetically valid species.

Genotypic identification is emerging as an alternative or complement to established phenotypic methods. The 16S rRNA gene is the most widely accepted gene used for bacterial classification and identification (35). Signature nucleotides of 16S rRNA genes allow classification even if a particular sequence has no match in the database, since otherwise unrecognizable isolates can be assigned to phylogenetic branches at the class, family, genus, or subgenus levels. This has contributed greatly to the discovery of new species of GPAC (8, 17, 21), and its variable regions have been used to design probes for detecting clinically significant GPAC (38, 39). Although the direct sequencing of amplified DNA from the 16S rRNA gene should allow unambiguous, definitive identification and provides information on the taxonomic relatedness of new species, its use for species identification based on public 16S rRNA databases is not without limitations (4, 25, 34). There are multiple problems with public database sequences, such as base errors, ambiguous base designations, and incomplete sequences, that may not be evident to users and will often lead to misidentification. Objective, clean 16S rRNA sequence data

* Corresponding author. Mailing address: 11301 Wilshire Blvd., Rm. E3-237, Bldg. 304, VA Medical Center West Los Angeles, Los Angeles, CA 90073. Phone: (310) 478-3711, ext. 49151. Fax: (310) 268-4458. E-mail: yulis1@yahoo.com.

TABLE 1. List of type strains and clinical isolates of GPAC species used in this study[a]

Species	Strain	No. of strains
Anaerococcus lactolyticus	CCUG 31351[T]	1
Anaerococcus octavius	CCUG 38493[T]	1
Anaerococcus prevotii	CCUG 41932[T]	1
Anaerococcus tetradius	CCUG 46590[T]	1
Anaerococcus vaginalis	CCUG 31349[T]	1
Finegoldia magna	CCUG 17636[T]	1
Micromonas micros	CCUG 46357[T]	1
Peptoniphilus assaccharolyticus	CCUG 9988[T]	1
Peptoniphilus harei	CCUG 38491[T]	1
Peptoniphilus indolicus	CCUG 17639[T]	1
Peptoniphilus ivorii	CCUG 38492[T]	1
Peptoniphilus lacrimalis	CCUG 31350[T]	1
Peptostreptococcus anaerobius	CCUG 7835[T]	1
Anaerococcus prevotii	Clinical isolates	20
Anaerococcus tetradius	Clinical isolates	10
Finegoldia magna	Clinical isolates	36
Micromonas micros	Clinical isolates	33
Peptoniphilus assaccharolyticus	Clinical isolates	30
Peptostreptococcus anaerobius	Clinical isolates	27

[a] Identification of clinical isolates was done by phenotypic tests.

for GPAC are important to determine the relationship of clinically relevant GPAC species.

In the present study, we determined the nearly complete 16S rRNA gene sequence data (>1,400 bp) for 13 type strains of established GPAC species. The sequence data obtained were compared to those in public sequence repositories such as GenBank. Based on the sequence data of the reference strains obtained in the present study, we evaluated 16S rRNA sequencing identification of GPAC by reidentifying a collection of 156 clinical isolates of GPAC that had previously been identified by phenotypic tests and represented the most commonly isolated clinical GPAC species.

MATERIALS AND METHODS

Bacterial strains and criteria for identification comparison. The bacterial strains used in the present study included 13 CCUG (Culture Collection, University of Göteborg) type strains of GPAC species and 156 isolates that were previously recovered from clinical specimens and identified by using standard phenotypic tests (Table 1). The isolates were chosen to represent six commonly isolated GPAC species based on phenotypic testing in our laboratory over a 5-year period. The strains were cultured overnight on brucella blood agar (Anaerobe Systems) at 37°C under an N_2 (86%), H_2 (7%), and CO_2 (7%) gas phase and were characterized biochemically by using a combination of conventional tests as described in the Wadsworth anaerobe manual (15) and the BD BBL Crystal Identification System (Becton Dickinson Microbiology Systems, Cockeysville, Md.).

Identifications were counted as correct if both methods provided the same answer. A discordant result was defined as a 16S rRNA sequence match with a species other than the species assigned by conventional identification.

16S rRNA gene sequencing. Genomic DNA was extracted and purified from cells in the mid-logarithmic-growth phase by using a QIAamp DNA Mini Kit (Qiagen, Inc., Chatsworth, Calif.). The 16S rRNA gene fragments were amplified as previously described (3). Briefly, two subregions of 16S rRNA gene were amplified by using two pairs of primers. The two subregions were defined as follows: region A was defined as 899-bp sequences between primer 8UA and 907B, and region B was defined as 711-bp sequences between primer 774A and 1485B. PCR was performed for 35 cycles of 30 s at 95°C, 30 s at 45°C, and 1 min at 72°C, with a final extension at 72°C for 5 min. PCR products were excised from a 1% agarose gel after electrophoresis and purified by using a QIAquick gel extraction kit (Qiagen) and were sequenced directly with a Biotech Diagnostics Big-Dye sequencing kit on an ABI 377 sequencer (Applied Biosystems, Foster City, Calif.).

Sequence data analysis. The sequencing data was analyzed as follows: (i) assembly of the reverse and forward sequences into a consensus sequence, (ii) editing of the consensus sequence to resolve discrepancies between the two strands by evaluation of the electropherograms, and (iii) comparison of the consensus sequences with GenBank sequences by using Ribosomal Database Project (RDP-II; Michigan State University, East Lansing) (18) and the basic local alignment search tool (BLAST) (1). Analysis of GPAC clinical isolates was performed by comparing the sequences against the sequences of type strains determined in the present study, as well as related sequences retrieved from GenBank. For an accurate determination of species similarities, all 5′ and 3′ ends were cut to identical positions along the gene, at *Escherichia coli* bp 49 to 1470. The newly determined sequences were aligned with their related sequences by using the program CLUSTAL W (13). The resulting multiple sequence alignment was corrected manually using the program GeneDoc (24). A phylogenetic tree was constructed by using the neighbor-joining algorithm PAUP 3.1.1 (D. L. Swofford, PAUP: phylogenetic analysis using parsimony [1993]).

RESULTS

16S rRNA genes of type strains of GPAC species. To demonstrate the quality and accuracy of results provided from a public database, we compared all type strain sequences determined in our laboratory to their corresponding GenBank sequences. Among 13 type strains we analyzed, only 4 strains had a "perfect" match (similarity, >99.5%) with sequences of their corresponding strains from GenBank as determined by using both BLAST and RDP-II (Sequence Match, version 2.7). The other nine strains showed low sequence similarities (≤98%) versus GenBank sequences because of abundant ambiguities, sequence gaps, and sequence errors (Table 2). For example, for the sequences of the type strains of *Anaerococcus prevotii*, *Micromonas micros*, and *Peptoniphilus asaccharolyticus* the best matches given by a BLAST search against GenBank were uncharacterized *Peptostreptococcus* sp. clone KL-59-7-12 (97.8%), *Peptostreptococcus* sp. oral clone FG 014 (99.6%), and *Peptostreptococcus* sp. strain S1 (97.9%), respectively, because the sequences of these type strains in the GenBank database were not of good quality. In the case of *Peptoniphilus indolicus*, the best sequence match given by both BLAST and RDP-II was a *Peptostreptococcus* sp. strain S1 instead of the GenBank sequence of *Peptoniphilus indolicus* type strain (GenBank accession no. D14147) and, indeed, the GenBank sequence of *Peptoniphilus indolicus* (D14147) was not even shown on the BLAST match list due to the low similarity between this two sequences. For the type strains of *A. lactolyticus*, *A. tetradius*, *A. vaginalis*, *Finegoldia magna*, and *Peptoniphilus lacrimalis*, although the best matches given by both BLAST and RDP-II were the sequences of their corresponding strains in GenBank, the sequence similarities between the sequences in the present study and those corresponding sequences in GenBank were very low (sequence similarities of ≤98% as determined by BLAST; similarity scores of ≤0.916 as determined by RDP-II).

Different search results between BLAST and RDP-II were also observed. For example, the highest matches for *A. prevotii* and *M. micros* sequences were uncultured *Peptostreptococcus* sp. clone KL-59-7-12 and *Peptostreptococcus* sp. oral clone, respectively, as determined by BLAST, but using RDP-II gave the best match against their corresponding species (*A. prevotii* and *M. micros*, respectively) with low similarity scores of 0.831 and 0.899, respectively (Table 2). For an accurate determination of species similarities, all 5′ and 3′ ends of the sequences were cut to identical positions along the gene, at *E. coli* bp 49 to 1450, and the sequences determined in the present study

TABLE 2. Comparison of 16S rRNA sequences of GPAC species obtained in this study with the sequences in GenBank

Species	Strain	Accession no.	Identity as determined by:		Comparison with sequence of corresponding strain in GenBank by CLUSTAL W alignment[a]	
			BLAST (%)	RDP-II (score)	Accession no.	Comments
Anaerococcus lactolyticus	CCUG 31351[T]	AF542233	*Anaerococcus lactolyticus* (98.0)	*Anaerococcus lactolyticus* (0.916)	D14154	14-Base divergence; 15-ambiguities
Anaerococcus prevotii	CCUG 41932[T]	AF542232	*Peptostreptococcus* sp. clone KL-59-7-12 (97.8)[b]	*Anaerococcus prevotii* (0.871)	D14139	37-Base divergence; 9-ambiguities
Anaerococcus tetradius	CCUG 46590[T]	AF542234	*Anaerococcus tetradius* (97.4)	*Anaerococcus tetradius* (0.820)	D14142	67-Base divergence; 8-ambiguities
Anaerococcus vaginalis	CCUG 31349[T]	AF542229	*Anaerococcus vaginalis* (97.9)	*Anaerococcus vaginalis* (0.905)	D14146	17-Base divergence; 10-ambiguities
Finegoldia magna	CCUG 17636[T]	AF542227	*Finegoldia magna* (98.0)	*Finegoldia magna* (0.850)	D14149	23-Base divergence; 18-ambiguities; 12-base gap
Micromonas micros	CCUG 46357[T]	AF542231	*Peptostreptococcus* sp. oral clone FG014 (99.6)[c]	*Micromonas micros* (0.889)	D14143	14-Base divergence; 9-ambiguities
Peptoniphilus asaccharolyticus	CCUG 9988[T]	AF542228	*Peptostreptococcus* sp. strain S1 (97.9)[d]	*Peptostreptococcus* sp. strain S1 (0.916)	D14138	14-Base divergence; 15-ambiguities
Peptoniphilus indolicus	CCUG 17639[T]	AY153431	*Peptostreptococcus* sp. strain S1 (99.2)[e]	*Peptostreptococcus* sp. strain S1 (0.960)[e]	D14147	141-Base divergence; 4-ambiguities
Peptoniphilus lacrimalis	CCUG 31350[T]	AF542230	*Peptoniphilus lacrimalis* (97.9)	*Peptoniphilus lacrimalis* (0.816)	D14141	46-Base divergence; 4-ambiguities
Anaerococcus octavius	CCUG 38493[T]		*Anaerococcus octavius* (100)	*Anaerococcus octavius* (1.000)	Y07841	
Peptoniphilus harei	CCUG 38491[T]		*Peptoniphilus harei* (100)	*Peptoniphilus harei* (1.000)	Y07839	
Peptoniphilus ivorii	CCUG 38492[T]		*Peptoniphilus ivorii* (100)	*Peptoniphilus ivorii* (1.000)	Y07840	
Peptostreptococcus anaerobius	CCUG 7835[T]		*Peptostreptococus anaerobius* (99.5)	*Peptostreptococcus anaerobius* (0.992)	L04168	8-Ambiguities

[a] All of the sequences were cut to identical positions along the 16S rRNA gene, corresponding to *E. coli* basepair positions 49 to 1450.
[b] The BLAST search showed only a 96.4% sequence similarity with *A. prevotii* (D14139) in GenBank.
[c] The BLAST search showed only a 98% sequence similarity with *M. micros* (D14143) in GenBank.
[d] The BLAST search showed only a 96.5% sequence similarity with *Peptoniphilus asaccharolyticus* (D14138) in GenBank.
[e] GenBank sequence (D14147) of *Peptostreptococcus indolicus* was not shown on the BLAST match list due to the low sequence similarity. Pairwise analysis indicated that it shared 98.0% similarity with the sequence of the *Peptoniphilus asaccharolyticus* type strain determined in this study.

and their corresponding sequences in GenBank were aligned and manually analyzed. The results are shown in Table 2.

Comparison between a conventional method and 16S rRNA sequencing for identification of GPAC clinical isolates. A total of 156 isolates, representing six clinically common GPAC species, were subjected to 16S rRNA sequence analysis. The breakdown of 156 clinical strains was as follows: 131 strains had a sequence with high similarity (>99%) to the type strains of an established species, 12 strains had a sequence similar to that of a *Peptostreptococcus* sp. oral clone in GenBank, and 13 strains had eight unique sequences that were distinct from any sequence of established species and from uncharacterized strains in the GenBank database. A comparison between sequence-based identification and phenotypic identification showed that 88 strains (56%) had concordant results between the original identification and the 16S ribosomal DNA sequencing identification, and the other 68 (44%) isolates had a molecular identification that was discordant with the original identification (Table 3). Two clinically significant species, *F. magna* (*n* = 36) and *M. micros* (*n* = 33), were consistently identified correctly by both phenotypic and genotypic methods. Totals of 56% (15 of 27) of *Peptostreptococcus anaerobius*, 5% (1 of 20) of *A. prevotii*, and 30% (3 of 10) of *A. tetradius* were

also identified correctly by both methods. Among the 88 discrepant isolates, 19 strains of *Peptoniphilus asaccharolyticus* and 19 strains of *A. prevotii* that were originally identified by phenotypic testing were reidentified as *Peptoniphilus hareii* by sequencing. Six isolates of *A. vaginalis* were misidentified as *A. tetradius* based on phenotypic tests, twelve isolates that were first identified as *Peptostreptococcus anaerobius* were determined to be 99% similar to an uncharacterized oral clone CK 035 and only 98.0% similar to *Peptostreptococcus anaerobius*. 16S rRNA sequence analysis indicated that 25 of 88 strains that were misidentified to species level by phenotypic tests may be novel species or subspecies since they had low sequence similarities against both the GenBank database and the sequences determined in the present study (Table 3). These 25 strains had nine unique sequences; these were assigned a phylogenetic position by building a phylogenetic tree with their related sequences (Fig. 1).

DISCUSSION

Accurate identification of bacterial isolates is an essential task of the clinical microbiology laboratory. It helps in evaluating the clinical significance of a positive culture and is im-

TABLE 3. Comparison of 16S rRNA sequencing identification with phenotypic identification

| Phenotypic identification (n) | 16S rRNA sequencing identification[a] (% identity) in comparison with: | | | | No. matched (%)[b] |
	Type strain sequence (% identity)	n	GenBank type strain sequence (% identity)	n	
Anaerococcus prevotii (20)	*Anaerococcus prevotii* (99.5)	1	Uncultured clone KL-59-7-12 (99.3)	1	1 (5)
	Peptoniphilus harei (>99.0)	18	*Peptoniphilus harei* (> 99.0)	18	
	Peptoniphilus asaccharolyticus (91.7)	1	***Peptostreptococcus* sp. (93.8)**	1	
Anaerococcus tetradius (10)	*Anaerococcus tetradius* (99.9)	3	*Anaerococcus tetradius* (97.4)	3	3 (30)
	Anaerococcus vaginalis (99.9)	6	*Anaerococcus vaginalis* (97.2)	6	
	Finegoldia magna (88.7)	1	***Finegoldia magna*** (86.9)	1	
Finegoldia magna (36)	*Finegoldia magna* (100)	36	*Finegoldia magna* (98.0)	36	36 (100)
Micromonas micros (33)	*Micromonas micros* (99.8)	33	*Peptostreptococcus* sp. oral clone FG 014 (99.9)	33	33 (100)
Peptostreptococcus anaerobius (27)	*Peptostreptococcus anaerobius* (99.8)	15	*Peptostreptococcus anaerobius* (99.8)	15	15 (55.6)
	Peptostreptococcus anaerobius (98.0)	12	*Peptostreptococcus* sp. oral clone CK 035 (99.9)	12	
Peptoniphilus asaccharolyticus (30)	*Peptoniphilus harei* (>99.0)	19	*Peptoniphilus harei* (>99.0)	19	0 (0)
	Peptoniphilus harei (97.2)	2	***Peptoniphilus harei*** (97.2)	2	
	Peptoniphilus harei (95.6)	3	***Peptoniphilus harei*** (95.6)	3	
	Peptoniphilus harei (93.2)	2	***Peptoniphilus harei*** (93.2)	2	
	Peptoniphilus ivorii (84.8)	1	***Peptoniphilus* genospecies 4** (89.2)	1	
	Micromonas micros (83.9)	1	***Micromonas micros*** (88.4)	1	
	Anaerococcus octavius (95.5)	2	***Anaerococcus octavius*** (95.5)	2	
Total					88 (56.4)

[a] Possible novel species are in boldface.
[b] That is, the number of matches between the phenotypic identification and the 16S rRNA sequencing identification based on type strain sequences determined in this study.

portant for predicting effective antibiotic therapy. Sequence-based analysis of the 16S rRNA gene represents a highly accurate and versatile method for bacterial classification and identification. The features of this molecular target, the universal distribution among bacteria (35), and the presence of species-specific variable regions make it a useful tool not only for bacterial phylogenetic analysis (36) but also for the detection and identification of clinical isolates (4, 5, 25–27, 30, 32–34). However, its use for identification is not without limitation. First, reliable 16S rRNA sequencing identification requires accurate sequence databases but, as indicated in Table 2, a great portion of the search results for GPAC type strains obtained from comparisons with GenBank submissions (the basis for both NCBI and RDP-II) were inaccurate (Table 2), with abundant ambiguities, sequence gaps, and errors. Among the 13 type strains we sequenced, 9 strains had low sequence similarities with their corresponding GenBank sequences, as determined by both BLAST and RDP-II. This is not surprising since the sequences of the majority of established GPAC species were submitted in the early 1990s (6, 8), when the methods used may not have been able to provide the quality sequences easily obtained today. These nine strains include some clinically significant GPAC species, such as *A. prevotii*, *A. tetradius*, *F. magna*, *M. micros*, and *Peptoniphilus asaccharolyticus*.

The type strain sequences in GenBank may provide mislead-ing identification results for clinical isolates (as shown in Table 3). For example, 33 clinical isolates of *M. micros* that were identified by 16S rRNA sequencing based on comparison to the sequence of the *M. micros* type strain obtained in the present study (sequence similarity of >99%) were given the highest match with the oral clone FG 014 (sequence similarity of >99%) as determined by BLAST search against GenBank, despite the fact that their type strain sequences were present in the database. Similarly, 6 strains of *A. vaginalis*, 36 strains of *F. magna*, and 3 strains of *A. tetradius* that were identified to the species level by showing high sequence similarities (>99%) with their corresponding type strain sequences determined in the present study were only given low sequence similarities (<97%) to corresponding GenBank sequences determined by both BLAST and RDP-II. Although there is no an accepted cutoff value of 16S rRNA sequence similarity for species definition (12, 31), it is apparent from the results of studies of numerous diverse taxa that the majority of recognized species that have been examined to date differ in their 16S rRNA sequence from related species of the same genus in at least 1% of the sequence positions—and typically more. Our 16S rRNA sequence data showed that the recognized GPAC species within a given genus have up to 8.0% or greater average divergence. The average species-species pair showed 12.9 and 8.0% sequence divergence within the genera *Peptoniphilus* and *Anaerococcus*, respectively, whereas the most similar pair (*Pep-*
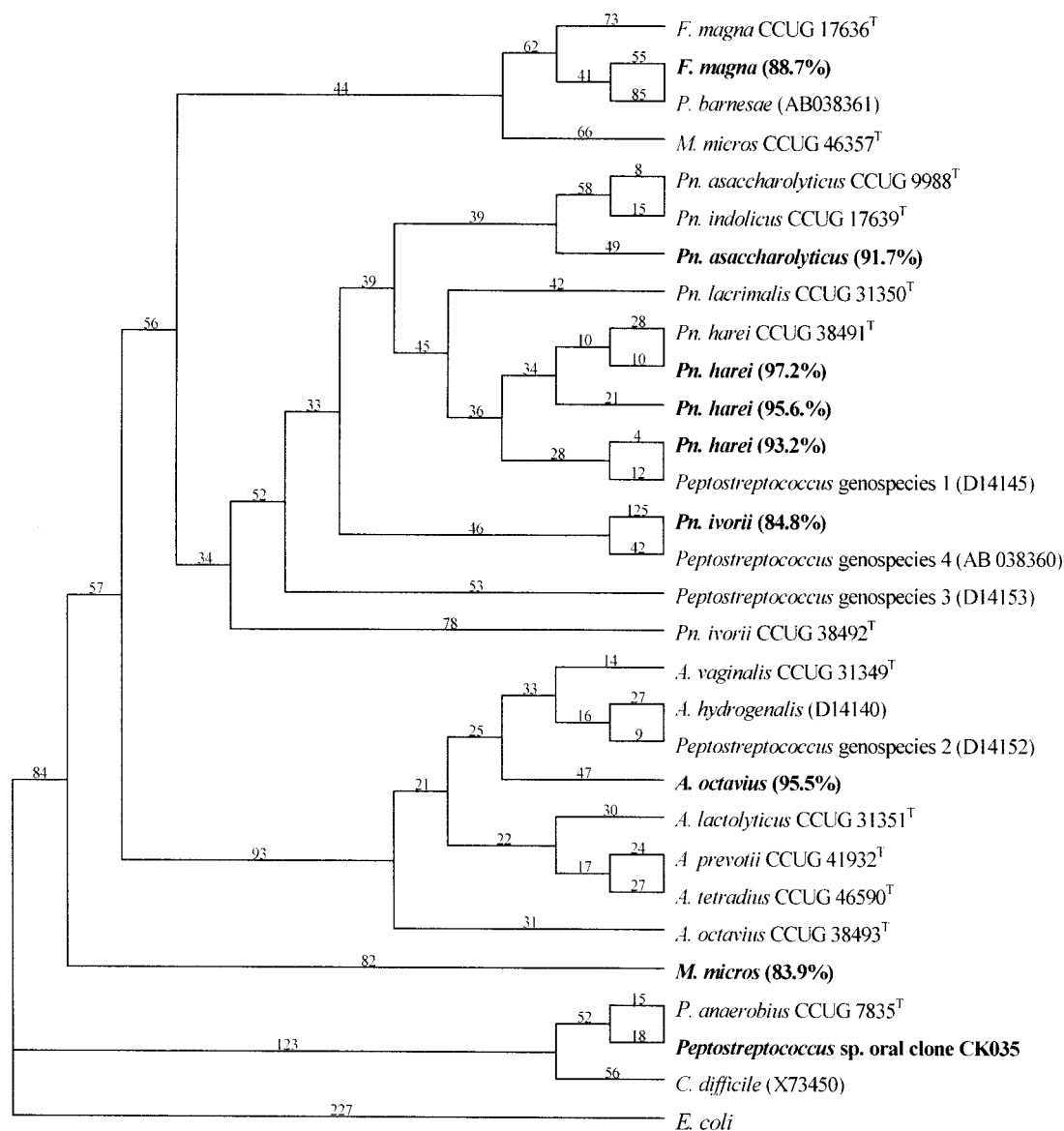
FIG. 1. Phylogenetic tree indicates the phylogenetic relationship of the nine novel species with their related established species, including one representative strain for each species. Sequences were determined in our laboratory unless indicated by a GenBank accession number. The tree was rooted by using *E. coli* as the outgroup sequence. Boldface type indicates possible novel species. The parenthetic percentage values indicate the 16S rRNA sequence similarities with corresponding species. Some of the genera in the phylogenetic tree are abbreviated as follows: *P.*, *Peptostreptococcus*; *F.*, *Finegoldia*; *M.*, *Micromonas*; *Pn.*, *Peptoniphilus*; *C.*, *Clostridium*.

toniphilus asaccharolyticus and *Peptoniphilus indolicus*) of the genus *Peptoniphilus* exhibited 98.2% sequence similarity. This indicates clearly that the recognized species of GPAC are typically separated by good evolutionary distances. Although the species of the genera *Peptoniphilus* and *Anaerococcus*, as currently constituted, may have to be reallocated to more than these two genera, in the present study we used 99% similarity (BLAST similarity of ≥99% and RDP similarity score of ≥0.99) as a suitable cutoff for identification at the species level. If an isolate showed a genetic difference of ≥1.00 and ≤2.00%, we reported it as closely related to its best match; thus, the 12 isolates that had a 98.0% sequence similarity with *Peptostreptococcus anaerobius* might be the same species or a subspecies

of *Peptostreptococcus anaerobius*. In cases in which the genetic difference is >2%, the isolate was reported as a unique isolate that may represent a novel taxon (Table 3). The reason we used only type strains was to eliminate any possible errors of species identification due to initial strain misidentification. We have been aware of the presence of sequences in GenBank that belong to strains of the same species sharing very low similarity (<85%) with each other and with their corresponding type strains. For example, the GenBank sequences of two *Peptoniphilus asaccharolyticus* strains, GIFU 7717 and GIFU 7946, had only 86.4 and 84.3% sequence similarities with the sequence of their type strain, respectively.

The purpose of performing sequence similarity searches

with the 16S rRNA sequence of type strains of GPAC was to evaluate the accuracy of results obtained by programs such as BLAST and RDP-II. We found that even when the same sequence was compared against the same database (GenBank) by using different programs (BLAST and RDP-II), different similarity results were obtained (Table 2), resulting in the assignment of different identities. This is because the similarity scores obtained depend on the length of the sequences under analysis and on the number of gaps introduced in the query sequence to optimize the similarity. Also, BLAST searches against all available sequences, whereas RDP-II acquires only select GenBank sequences and incorporates them into their own database against which searches are made. In the present study, for accurate identification, we felt that all of the sequences included in the similarity search should be cut to the same length and, in addition to similarity search through internet available programs such as BLAST and RDP-II, the sequences should also be analyzed by multiple sequence alignment followed by manual correction.

The 16S rRNA sequencing method for identification of clinical isolates of GPAC based on accurate type strain sequences determined in the present study was evaluated by blindly reidentifying a collection of 156 GPAC isolates previously identified by phenotypic testing. The 16S rRNA sequencing identification proved to be more accurate than the phenotypic identification. We found this approach to be efficient in the majority of cases, with 92% (143 of 156) of isolates being identified to the species level, in contrast to 56.4% being identified by phenotypic identification, which is biased by errors and the variability of character expression. By 16S rRNA sequence analysis, even the 13 isolates that could not be identified to the species level could be assigned to a phylogenetic position (Fig. 1). Comparison between the 16S rRNA sequencing identification and phenotypic identification showed that two clinically significant species, *F. magna* ($n = 36$) and *M. micros* ($n = 33$), were consistently identified correctly by both methods. However, all *Peptoniphilus harei* isolates were misidentified as *Peptoniphilus asaccharolyticus* or *A. prevotii* by biochemical tests. *Peptoniphilus harei* was recently distinguished from *Peptoniphilus asaccharolyticus* by Murdoch and Mitchelmore (22); it resembles *Peptoniphilus asaccharolyticus* biochemically. Phenotypically, it is differentiated from *Peptoniphilus asaccharolyticus* by cell and colony morphology, which may be very subjective. Our results showed that sequence-based identification was better able to distinguish heterogeneous *A. prevotii* isolates (9, 17) than phenotypic identification. The observations of the present study indicate a greater clinical importance of *Peptoniphilus harei* than would be indicated by phenotypic tests. Although 16S rRNA sequencing provides the advantage of accuracy for GPAC identification compared to phenotypic identification, cost is a critical issue in the evaluation of 16S rRNA sequence-based analysis as a diagnostic tool. The initial cost of equipment can be recovered quickly with savings in personnel, time, and ultimately in health care costs. Furthermore, driven in part by the technology underlying the human and microbial genome projects, sequencing costs will probably continue their rapid trend downward, bringing this technology within the reach of many microbiology laboratories.

In summary, 16S rRNA sequencing proved to be an accurate identification method for GPAC species. It not only allows proper identification of isolates but also rapid recognition and classification of previously undescribed organisms. More efforts should be made to complete 16S rRNA databases with high-quality sequences.

REFERENCES

1. **Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, B. A. Rapp, and D. L. Wheeler.** 2002. GenBank. Nucleic Acids Res. **30:**17–20.
2. **Brook, I.** 1988. Recovery of anaerobic bacteria from clinical specimens in 12 years at two military hospitals. J. Clin. Microbiol. **26:**1181–1188.
3. **Brosius, J., M. L. Palmer, P. J. Kennedy, and H. F. Noller.** 1978. Complete nucleotide sequence of a 16S ribosomal RNA gene from *Escherichia coli.* Proc. Natl. Acad. Sci. USA **75:**4801–4805.
4. **Cloud, J. L., H. Neal, R. Rosenberry, C. Y. Turenne, M. Jama, D. R. Hillyard, and K. C. Carroll.** 2002. Identification of *Mycobacterium* spp. by using a commercial 16S ribosomal DNA sequencing kit and additional sequencing libraries. J. Clin. Microbiol. **40:**400–406.
5. **Drancourt, M., C. Bollet, A. Carlioz, R. Martelin, J. P. Gayral, and D. Raoult.** 2000. 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. J. Clin. Microbiol. **38:**3623–3630.
6. **Ezaki, T., Y. Kawamura, N. Li, Z. Y. Li, L. Zhao, S. Shu.** 2001. Proposal of the genera *Anaerococcus* gen. nov., *Peptoniphilus* gen. nov. and *Gallicola* gen. nov. for members of the genus *Peptostreptococcus.* Int. J. Syst. Evol. Microbiol. **51:**1521–1528.
7. **Ezaki, T., and E. Yabuuchi.** 1985. Oligopeptidase activity of gram-positive anaerobic cocci used for rapid identification. J. Gen. Appl. Microbiol. **31:**255–266.
8. **Ezaki, T., S., L. Liu, Y. Hashimoto, and E. Yabuuchi.** 1990. *Peptostreptococcus hydrogenalis* sp. nov. from human fecal and vaginal flora. Int. J. Syst. Bacteriol. **40:**305–306.
9. **Ezaki, T., N. Yamamoto, K. Ninomiya, S. Suzuki, and E. Yabuuchi.** 1983. Transfer of *Peptococcus indolicus, Peptococcus asaccharolyticus, Peptococcus prevotii* and *Peptococcus magnus* to the genus *Peptostreptococcus* and proposal of *Peptostreptococcus tetradius* sp. nov. Int. J. Syst. Bacteriol. **33:**683–698.
10. **Finegold, S. M.** 1977. Anaerobic bacteria in human disease. Academic Press, Inc., New York, N.Y.
11. **Finegold, S. M.** 1995. Anaerobic infections in humans: an overview. Anaerobe **1:**3–9.
12. **Fox, G. E., J. D. Wisotzkey, and P. Jurtshuk, Jr.** 1992. How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int. J. Syst. Bacteriol. **42:**166–170.
13. **Higgins, D., J. Thompson, T. Gibson, J. D. Thompson, D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
14. **Holland, J. W., E. O. Hill, and W. A. Altemeier.** 1977. Numbers and types of anaerobic bacteria isolated from clinical specimens since 1960. J. Clin. Microbiol. **5:**20–25.
15. **Jousimies-Somer, H., P. Summanen, D. M. Citron, E. J. Baron, H. M. Wexler, and S. M. Finegold.** 2002. Wadsworth-KTL anaerobic bacteriology manual, 6th ed. Star Publishing, Belmont, Calif.
16. **Karachewski, N. O., E. L. Busch, and C. L. Wells.** 1985. Comparison of PRAS II, RapID ANA, and API 20A systems for identification of anaerobic bacteria. J. Clin. Microbiol. **21:**122–126.
17. **Li, N., Y. Hashimoto, S. Adnan, H. Miura, H. Yamamoto, and T. Ezaki.** 1992. Three new species of the genus *Peptostreptococcus* isolated from humans: *Peptostreptococcus vaginalis* sp. nov., *Peptostreptococcus lacrimalis* sp. nov., and *Peptostreptococcus lactolyticus* sp. nov. Int. J. Syst. Bacteriol. **42:**602–605.
18. **Maidak, B. L., J. R. Cole, T. G. Lilburn, C. T. Parker, Jr., P. R. Saxman, R. J. Farris, G. M. Garrity, G. J. Olsen, T. M. Schmidt, and J. M. Tiedje.** 2001. The RDP-II (Ribosomal Database Project). Nucleic Acids Res. **29:**173–174.
19. **Murdoch, D. A., I. J. Mitchelmore, and S. Tabaqchali.** 1994. The clinical importance of gram-positive anaerobic cocci isolated at St Bartholomew's Hospital, London, in 1987. J. Med. Microbiol. **41:**36–44.
20. **Murdoch, D. A.** 1998. Gram-positive anaerobic cocci. Clin. Microbiol. Rev. **11:**81–120.
21. **Murdoch, D. A., M. D. Collins, A. Willems, J. M. Hardie, K. A. Young, and J. T. Magee.** 1997. Description of three new species of the genus *Peptostreptococcus* from human clinical specimens: *Peptostreptococcus harei* sp. nov., *Peptostreptococcus ivorii* sp. nov. and *Peptostreptococcus octavius* sp. nov. Int. J. Syst. Bacteriol. **47:**781–787.
22. **Murdoch, D. A., and I. J. Mitchelmore.** 1991. The Laboratory identification of gram-positive anaerobic cocci. J. Med. Microbiol. **34:**295–308.
23. **Murray, P. R., C. J. Weber, and A. C. Niles.** 1985. Comparative evaluation of three identification schemes for anaerobes. J. Clin. Microbiol. **22:**52–55.

24. **Nicholas, K. B., H. B. Nicholas, Jr., and D. W. Deerfield II.** 1997. GeneDoc: analysis and visualization of genetic variation. EMBNEW News **4:**14.

25. **Patel, J. B., D. G. Leonard, X. Pan, J. M. Musser, R. E. Berman, and I. Nachamkin.** 2000. Sequence-based identification of *Mycobacterium* species using the MicroSeq 500 16S rDNA bacterial identification system. J. Clin. Microbiol. **38:**246–251.

26. **Patel, R., K., E. Piper, M. S. Rouse, J. M. Steckelberg, J. R. Uhl, P. Kohner, M. K. Hopkins, F. R. Cockerill III, and B. C. Kline.** 1998. Determination of 16S rRNA sequences of enterococci and application to species identification of nonmotile *Enterococcus gallinarum* isolates. J. Clin. Microbiol. **36:**3399–3407.

27. **Rogall, T., T. Flohr, and E. C. Bottger.** 1990. Differentiation of *Mycobacterium* species by direct sequencing of amplified DNA. J. Gen. Microbiol. **36:**1915–1920.

28. **Rogosa, M.** 1974. Family III: *Peptococcaceae*, p. 517–527. *In* R. E. Buchanan and N. E. Gibbons (ed.), Bergey's manual of determinative bacteriology, 8th ed. The Williams & Wilkins Co., Baltimore, Md.

29. **Song, Y. L., C. X. Liu, D. Molitoris, T. J. Tomzynski, M. McTeague, E. Read, and S. M. Finegold.** 2002. Use of 16S-23S rRNA spacer-region (SR)-PCR for identification of intestinal clostridia. Syst. Appl. Microbiol. **4:**528–535.

30. **Springer, B., L. Stockman, K. Teschner, G. D. Roberts, and E. C. Bottger.** 1996. Two-laboratory collaborative study on identification of mycobacteria: molecular versus phenotypic methods. J. Clin. Microbiol. **34:**296–303.

31. **Stackebrandt, E., and B. M. Goebel.** 1994. A place for DNA-DNA hybridization reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. **44:**846–849.

32. **Tang, Y. W., A. Von Graevenitz, M. G. Waddington, M. K. Hopkins, D. H. Smith, H. Li, C. P. Kolbert, S. O. Montgomery, and D. H. Persing.** 2000. Identification of coryneform bacterial isolates by ribosomal DNA sequence analysis. J. Clin. Microbiol. **38:**1676–1678.

33. **Tang, Y. W., N. M. Ellis, M. K. Hopkins, D. H. Smith, D. E. Dodge, and D. H. Persing.** 1998. Comparison of phenotypic and genotypic techniques for identification of unusual aerobic pathogenic gram-negative bacilli. J. Clin. Microbiol. **36:**3674–3679.

34. **Turenne, C. Y., L. Tschetter, J. Wolfe, and A. Kabani.** 2001. Necessity of quality-controlled 16S rRNA gene sequence databases: identifying nontuberculous *Mycobacterium* species. J. Clin. Microbiol. **39:**3637–3648.

35. **Weisburg, W. G., S. M. Barns, D. A. Pelletier, and D. J. Lane.** 1991. 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. **173:**697–703.

36. **Woese, C. R., O. Kandler, and M. I., Wheelis.** 1990. Towards a natural system of organism: proposal for the domains *Archaea*, *Bacteria*, and *Eukarya*. Proc. Natl. Acad. Sci. USA **87:**4576–4579.

37. **Wren, M. W. D., A. W. F. Baldwin, C. P. Eldon, and P. J. Sanderson.** 1977. The anaerobic culture of clinical specimens: a 14-month study. J. Med. Microbiol. **10:**49–61.

38. **Yasui, S.** 1989. Development and clinical application of DNA probe specific for *Peptostreptococcus micros*. Bull. Tokyo Med. Dent. Univ. **36:**49–62.

39. **Yasui, S., H. S. Reynolds, J. J. Zambon, R. J. Genco, and T. V. Potts.** 1989. Development of specific DNA probe for the identification of *Peptostreptococcus anaerobius*. J. Dent. Res. **68:**256.